



Economic and Social Council

Distr.: General
19 March 2014

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-second plenary session

Paris, 9-11 April 2014

Item 8 (a) of the provisional agenda

Reports on the work of the Conference of European Statisticians, its Bureau and Teams of Specialists

Report of the third meeting of the 2013/2014 Bureau of the Conference of European Statisticians

Note by the Secretariat

Addendum

Summary

The present report is submitted in compliance with the Rules Governing the Work of the Conference and its Bureau (ECE/CES/2007/8), adopted by the Conference of European Statisticians at its fifty-third plenary session. Referring to the Rules, the Bureau acts as a steering body of the Conference. The report includes decisions taken by the Bureau at its 28-29 January 2014 meeting on the work carried out under the auspices of the Conference.

GE.14-21377



★ 1 4 2 1 3 7 7 ★

Please recycle



Contents

	<i>Page</i>
I. Introduction	3
II. In-depth reviews of selected statistical areas	3
A. Measuring social protection.....	3
B. Leading, composite and sentiment indicators and official statistics	4
C. Follow-up to the in-depth review of political and other community activities and volunteering ..	5
D. Follow-up to the in-depth review of entrepreneurship.....	5
E. Selection of topics for in-depth reviews in 2014-2015	6
F. Updated template for in-depth review papers.....	6
III. UNECE statistical work and CES teams of specialists.....	7
A. Outcome of the seminar on poverty measurement.....	7
B. Time-use surveys	7
C. Migration statistics.....	8
D. Measuring human capital.....	9
E. Measuring sustainable development and implementation of SEEA	9
F. Draft guide to measuring global production	9
G. Outcomes of the projects working under the High-Level Group for the Modernisation of Statistical Production and Services in 2013.....	10
IV. Preparation for the CES 2014 plenary session.....	11
A. Organisation of the CES 2014 seminar "Migration statistics".....	11
B. Organisation of the CES 2014 seminar "What is the value of official statistics and how do we communicate that value?"	11
C. Recommendations on climate change related statistics	12
D. Draft agenda and timetable for the 2014 plenary session	13
V. UNECE programme of work documents.....	13
VI. Organisation of CES work.....	13
A. Updated guidelines for organising the CES seminars.....	13
B. Governing rules of CES.....	13
VII. Presentation of ILO activities related to statistics.....	14
VIII. Dates and venue of the next meetings.....	15

I. Introduction

1. The third meeting of the 2013/2014 Bureau was held in Geneva on 28-29 January 2014. The following members of the Bureau attended: Mr. W. Smith (Chairman), Mr. B. Aydemir (Turkey), Ms. M. Bruun (Finland), Ms. J. Matheson (the United Kingdom), Mr. K. Pesendorfer (Austria) and Mr. A. M. Ortega (representing Mr. Eduardo Sojo Garza-Aldape, Mexico). The following permanent observers also attended: Mr. L. M. Ducharme (IMF), Ms. M. Durand, (OECD), Ms. C. Sá (representing Mr. W. Radermacher, Eurostat), Mr. A. Kosarev (representing Mr. Sokolin, CIS-STAT) and Ms. L. Bratanova (UNECE). Mr. S. Sperlich (University of Geneva) and Mr. P. Dembinski (Observatoire de la Finance, Genève) attended for agenda item 2 (b), Mr. R. Smith for agenda item 4 (c) (via an audioconference), and Mr. R. Diez de Medina for agenda item 7.
2. The following persons assisted the members of the Bureau: Ms. A. L. MacMillan (Canada), Ms. S. Onur (Turkey) and Ms. M. J. Santos (Eurostat). Ms. T. Luige of UNECE served as Secretary of the meeting.
3. The following staff of the UNECE Statistical Division also attended: Mr. C. Boldsen, Ms. T. Dimova, Ms. A. Peltola, Mr. S. Vale, Mr. P. Valente and Mr. A. Vikat.

II. In-depth reviews of selected statistical areas

A. Measuring social protection

Documentation: ECE/CES/BUR/2014/JAN/2 by UNECE

4. The CES Bureau conducted an in-depth review of measuring social protection based on a paper by the Secretariat. The following issues were raised in the discussion:
 - (a) Measuring social protection is a complex area of high policy importance. It covers a broad range of issues and involves different institutions at national and international level, many of these from outside official statistics;
 - (b) The differences in measuring social protection stem from different uses of the data, and it may not be possible to impose universal standards for all purposes. However, a certain level of harmonization is needed to facilitate international comparability and ensure that the data are based on sound methodology. There is currently a momentum to define basic standards in this area. The Social Protection Inter-agency Cooperation Board (SPIAC-B) is working towards harmonization in this field;
 - (c) Any possible future work on this topic under CES should add value to the existing initiatives;
 - (d) The suggestions for future work included organizing a CES seminar to discuss the issues and challenges in this area. The following potential work areas were mentioned: developing a small set of internationally comparable indicators on social protection, considering the recipient side of social protection, the parts of society not covered by social protection, use of administrative data, and funding of the social security. It was noted that some aspects of measuring social protection could be considered within the Task Force on poverty.

Conclusions

5. The Bureau recognised the need for more harmonisation in measuring social protection. The Secretariat, in consultation with OECD, World Bank and ILO will inform the Bureau about the outcome of the meeting of the SPIAC-B group in May 2014. On this basis, the Bureau will decide about possible further work related to measuring social protection at its October 2014 meeting.

B. Leading, composite and sentiment indicators and official statistics

Documentation: ECE/CES/BUR/2014/JAN/3 by UNECE

6. The Bureau conducted an in-depth review of leading, composite and sentiment indicators and official statistics based on a paper by the Secretariat.

7. Mr. Dembinski (Observatoire de la Finance, Geneva, University of Fribourg) had contacted the UNECE Secretariat before the CES Bureau meeting with a proposal to cooperate on a research project on composite indicators. The aim of the project is to review a number of composite indicators and analyse their conceptual meaningfulness, statistical robustness and role in public debates. Mr. Sperlich (University of Geneva) and Mr. Dembinski were invited to the Bureau meeting to present the proposal.

8. The following issues were raised in the discussion:

(a) Strong demand exists for leading, composite and sentiment indicators. If statistical offices do not use their data and expertise to produce these indicators, other organizations may compile them with lower quality;

(b) On one hand, official statisticians are very cautious about leading, composite and sentiment indicators. Some statistical offices do not consider that producing such indicators provides a competitive advantage and is a good use of the limited resources. On the other hand, many national statistical offices and international organizations (such as OECD) are already producing leading, composite and sentiment indicators. Official statistics provide the quality criteria to ensure that the indicators are based on a sound methodology and can guide users on how to interpret the indicators. Producing these kind of indicators can improve users' perception of the relevance and value of official statistics;

(c) The area lacks international coordination and a systematic approach. There is a need for discussion at international level to achieve a common understanding of the role of statistical offices in this area. The in-depth review paper presents good proposals for further work. Organising a workshop or seminar on this topic could be considered;

(d) Collaboration with researchers is important. The research project proposed by the Swiss universities was considered useful. It was suggested that the project could focus on a few policy areas instead of covering the whole range. The outcomes of the project would be a valuable contribution to the discussion;

(e) It would be useful to develop a taxonomy of leading, composite and sentiment indicators. Various indicators are produced in countries and there is room for harmonization to enable better international comparability and reliability;

(f) The challenges related to leading and composite indicators are different from the sentiment indicators. There is a link between leading indicators and Big Data. Compiling composite indicators may require value judgments concerning the weighting of different components. This is not in line with the Fundamental Principles of official statistics;

(g) Providing a quality checklist for users might be a practical approach;

(h) The Active Ageing Index produced by UNECE will be added to the indicators described in the paper;

(i) Eurostat provided detailed comments in writing focusing on the need for a coherent classification and giving further information on on-going activities.

Conclusions

9. The Secretariat will update the in-depth review paper to take into account comments by the Bureau. The topic will be discussed at the CES plenary session in April 2014 to seek input from the CES members. The Secretariat will prepare a proposal for further work, based on the CES discussion, for the October 2014 Bureau meeting.

10. The Secretariat will follow up on the Swiss research project on composite indicators.

C. Follow-up to the in-depth review of political and other community activities and volunteering

Documentation: ECE/CES/BUR/2014/JAN/4 by UNECE

11. The Bureau discussed possible follow-up to the in-depth review of measuring political and other community activities, including volunteering, which was conducted by the Bureau in February 2013. The following issues were raised in the discussion:

(a) The potential CES seminar topic should have a clear focus, it is not recommended to broaden the topic to cover social capital, wellbeing, etc.;

(b) As a first step, a small group of countries could draft a possible road map in this area. It will be important to identify a country to lead the group.

Conclusions

12. The Secretariat will consult with a small group of countries/organizations (such as those who were interested in a CES seminar on this topic) to develop a proposal on the way forward in this area, and report to the October 2014 Bureau meeting.

D. Follow-up to the in-depth review of entrepreneurship

Documentation: ECE/CES/BUR/2014/JAN/5 by OECD, Eurostat and UNECE

13. The Bureau discussed follow-up to the in-depth review of entrepreneurship statistics which was conducted by the Bureau in October 2013. The following issues were raised in the discussion:

(a) The in-depth review contributed to a renewed momentum in entrepreneurship statistics. Issues related to entrepreneurship are currently tackled by the CES Task Force on Statistical Business Registers and the joint Eurostat/OECD Entrepreneurship Indicators Programme (EIP). The next meeting of EIP in May 2014 will be open to non-EU/OECD member countries;

(b) The Bureau acknowledged the work of the Task Force on Statistical Business Registers as a model of effective inter-organizational collaboration, and noted that the upcoming guidelines on business registers will be critical for compiling comparable entrepreneurship statistics. The Task Force was encouraged to follow the outcomes of the EIP work related to entrepreneurship indicators;

(c) Concern of possible duplication of work among the groups working on business registers was raised, namely the Wiesbaden Group on Business Registers and the joint UNECE/Eurostat/OECD Expert Group on Business registers. There are synergies

between the groups: they meet on alternate years and have a largely overlapping membership. The CES Task Force on Statistical Business Registers was considered to have a clear mandate with concrete expected outcomes (the group will be closed after submitting the Guidelines on Statistical Business Registers to the CES plenary session for approval in 2015).

Conclusions

14. The Bureau asked Eurostat and OECD to prepare, in collaboration with the Secretariat, a proposal for possible future activities on entrepreneurship statistics, taking into account the outcome of the meeting of the Eurostat/OECD Entrepreneurship Indicators Programme in May 2014.

15. The Secretariat will clarify the mandate and role of the Wiesbaden Group in consultation with the current chair of the Steering Group of the Wiesbaden Group and the UN Statistics Division in New York since the group reports to the UNSC. The Bureau will come back to the issue in October 2014.

E. Selection of topics for in-depth reviews in 2014-2015

Documentation: ECE/CES/BUR/2014/JAN/6 by UNECE

Conclusions

16. The following topics were selected for the next in-depth reviews, including the countries responsible for preparing the papers to provide basis for discussion:

(a) Demographic projections (Canada and the United Kingdom, review in October 2014);

(b) Measuring extreme events and disasters (Mexico and possibly New Zealand, review in October 2014);

(c) Process-oriented approach to statistical production (Turkey, review in February 2015);

(d) Labour movements and globalisation (Austria, review in February 2015).

17. The Secretariat will invite Statistics New Zealand to contribute to the in-depth review on measuring extreme events and disasters as a co-author with Mexico.

18. The Bureau selected the leading, composite and sentiment indicators for discussion at the 2014 CES plenary session to obtain input from all CES members.

F. Updated template for in-depth review papers

Documentation: ECE/CES/BUR/2014/JAN/7 by UNECE

Conclusions

19. The Bureau approved the updated template for in-depth review papers (document ECE/CES/BUR/2014/JAN/7).

III. United Nations Economic Commission for Europe statistical work and teams of specialists

A. Outcome of the seminar on poverty measurement

Documentation: ECE/CES/BUR/2014/JAN/8 by UNECE

20. The Bureau discussed the outcome of the seminar on the Way forward in poverty measurement held on 2-4 December 2013, and reviewed the proposed terms of reference for a Task Force on Poverty Measurement. The following issues were raised in the discussion:

(a) In view of the Post-2015 agenda, the Task Force should consider intra-household poverty (poverty among children, women and elderly) and explore how other concepts, such as inequality and social exclusion, relate to poverty;

(b) The work of the Task Force will be challenging due to the proliferation of poverty indicators. Defining relative and absolute poverty may be controversial and will, thus, require wide consultations. The Task Force should work in stages and build upon the lessons learned from previous initiatives, such as the different approaches adopted by the World Bank, OECD, Rio city group and the Oxford Poverty and Human Development Initiative;

(c) The Task Force should pay special attention to issues related to communicating information about poverty to policy makers and the general public;

(d) The next seminar on poverty could take place in spring 2015 allowing to take stock of country practices and available indicators on poverty, and discuss the Task Force's initial findings;

(e) The Task Force could invite renowned poverty measurement experts from the research community to join the group.

Conclusions

21. The Bureau supported the work and adopted the terms of reference of the Task Force on Poverty Measurement ECE/CES/BUR/2014/JAN/8), subject to the above mentioned changes. The Secretariat will ensure that the issues raised in the discussion are reflected in the work plan and schedule of the Task Force. Eurostat expressed interest to participate in the Task Force.

22. The next seminar on poverty and meeting of the task force are tentatively scheduled in spring 2015.

B. Time-use surveys

Documentation: ECE/CES/BUR/2014/JAN/9 by UNECE

23. The Bureau discussed possible follow-up work on time-use surveys after endorsing the Guidelines for harmonizing time-use surveys in October 2013. The Secretariat prepared a proposal for further work based on the issues identified in the Guidelines. The following points were raised in the discussion:

(a) The planned web-based collection of methodological material should be done in collaboration with Eurostat's data collection for the 2010 round of harmonized time-use statistics. The resulting pool of countries' methodologies would also serve for updating the

EU guidelines on time-use surveys. The Secretariat could consider preparing a synthesis report based on the collected material;

(b) There is a momentum to define unpaid household work and explore ways to measure it. Attaching a monetary value to unpaid work shows the importance that society attributes to this type of work as a support to the economy. Valuing unpaid work was discussed in the context of the revision of the 2008 SNA and measuring sustainable development. It was also one of Stiglitz commission's recommendations. Time-use surveys can be good sources of information for these purposes;

(c) A possible CES Task Force work in this area should be kept closely connected with EU-SILC and the ILO and OECD activities on voluntary work.

Conclusions

24. The Bureau supported the proposed activities for further work, namely (a) collecting methodological material on time-use surveys and making it available on the UNECE website, and (b) consulting with CES members a proposal to set up a Task Force for developing guidelines on valuing unpaid household work. Finland, CIS-STAT and Eurostat expressed their interest to participate in the Task Force. The Secretariat will report back to the Bureau in October 2014.

C. Migration statistics

Documentation: ECE/CES/BUR/2014/JAN/10 and 11 by the Steering Group and Task Force

25. The Bureau reviewed the progress of work of the Steering Group on Migration Statistics and the Task Force on Measurement of the Socio-economic Conditions of Migrants, including the proposals to extend the mandates of the groups. The following issues were raised in the discussion:

(a) The Bureau considered the work on migration statistics very important for the UNECE region. Given this, the Steering Group should be more active in identifying emerging issues and ensuring that the work is focused and progresses effectively;

(b) The role of the Steering Group should be strengthened in light of the outcome of the 2014 CES seminar on migration statistics and based on an evaluation of the impact of previous Task Forces on migration statistics;

(c) Methodological work should be undertaken to study the longitudinal aspects of migration, integration of second-generation migrants, migrant trajectories and emigration;

(d) The Task Force on Measurement of the Socio-economic Conditions of Migrants has taken a good initiative to separate population of interest into two groups – the foreign-born and the second-generation descendants. Given the many relevant socio-economic indicators, the Task Force should keep its practical focus.

Conclusions

26. The Bureau extended the mandate of the Task Force on Measurement of the Socio-economic Conditions of Migrants for one year.

27. The Secretariat will revise the terms of reference of the Steering Group on migration statistics to take into account the comments by the Bureau and the outcome of the CES seminar on migration statistics in April 2014. Mexico expressed its interest in participating in the Steering Group. The Secretariat will report back to the Bureau in October 2014.

D. Measuring human capital

Documentation: ECE/CES/BUR/2014/JAN/12 by UNECE

Conclusions

28. The Bureau reviewed the progress of work of the Task Force on Measuring Human Capital, and asked the Task Force to proceed in line with its terms of reference.

E. Measuring sustainable development and implementation of SEEA

Documentation: ECE/CES/BUR/2014/JAN/13 by UNECE

29. The Bureau discussed the follow-up to the CES 2013 seminar on challenges in measuring sustainable development and implementing the System of Environmental-Economic Accounting (SEEA) based on a paper prepared by the Secretariat. The following issues were raised in the discussion:

(a) In addition to the UNECE pilot testing of the indicators with eight volunteering countries (Australia, Italy, Kazakhstan, Mexico, Russian Federation, Slovenia, Turkey and Ukraine), OECD plans to include in their 2015-16 work plan the testing of indicators from the CES Recommendations on measuring sustainable development in OECD countries with the aim of using them for monitoring and benchmarking the comparative performance of countries. The testing will be done in close collaboration with UNECE;

(b) Eurostat welcomes the initiative to raise awareness about the CES Recommendations and supports the work on adjusting the indicators from the CES Recommendations to the Sustainable Development Goals;

(c) Many countries with developing statistical systems need capacity building in the implementation of SEEA.

Conclusions

30. The Bureau supported the follow-up activities to the CES 2013 seminar as proposed in document ECE/CES/BUR/2014/JAN/13, namely:

(a) Pilot testing of the CES Recommendations on measuring sustainable development: mapping the indicators of the CES Recommendations with national indicators, exploring data availability, and compiling case studies on the experiences in using the CES Recommendations in Italy, Mexico, Slovenia and Ukraine;

(b) Preparing stocktaking reports on measuring transboundary impacts in the context of sustainable development, measuring sustainable development at different scales, and linking subjective and objective indicators;

(c) Organising a joint UNECE/OECD seminar on the implementation of SEEA.

F. Draft Guide to measuring global production

Documentation: ECE/CES/BUR/2014/JAN/14 by the Task Force

31. The Guide to measuring global production was consulted with Bureau members prior to the meeting. All eight replies received were of the opinion that the Guide is useful

and ready to be sent for electronic consultation to all CES members. The following issues were raised in the written replies:

(a) The Task Force has accomplished important work. The Guide should be circulated beyond CES members to all national statistical offices and relevant international organizations in line with the UN Statistical Commission's procedures for providing guidance on SNA recommendations. The Secretariat will cooperate with UNSD to organize a global consultation of the Guide;

(b) The Guide should be updated in 5 to 7 years from now to reflect the new country circumstances, lessons learned from its implementation and to identify new case studies;

(c) Recording global production arrangements in national and international accounts presents numerous and complex challenges. Countries need a forum to discuss implementation challenges, share experience and coordinate the development work. The Steering Group on National Accounts should develop a proposal how the Joint UNECE/Eurostat/OECD Group of Experts on National Accounts could respond to this need.

Conclusions

32. The Secretariat will circulate the draft Guide to all CES members for electronic consultation in view of finalizing it in the course of 2014. The Guide is planned to be presented to the 2015 CES plenary session for endorsement.

33. The Steering Group on National Accounts will develop a proposal for establishing a forum on global production, possibly in the form of an expert meeting or a special session where ongoing issues related to global production could be discussed. The proposal should be submitted to the October 2014 Bureau meeting.

G. Outcomes of the projects working under the High-level Group for the modernisation of statistical production and services in 2013

Documentation: ECE/CES/BUR/2014/JAN/15 by the High-level Group

34. The Bureau reviewed the progress of work and the outcomes of two international collaboration projects carried out under the High-level Group (HLG) during 2013. The first project concerned the development of a Common Statistical Production Architecture, and the second one the development and enhancement of the Frameworks and Standards for Statistical Modernisation.

35. The following issues were raised in the discussion:

(a) The project outputs are very useful, represent good value for money, and are already being used in several organisations;

(b) Two projects to be completed by the end of 2014 are under way, namely on the implementation of the Common Statistical Production Architecture and the use of "Big Data" for official statistics. More components that are compliant with the Common Statistical Production Architecture will become available by the end of 2014. This will make it easier to promote the project outputs beyond participating countries and ensure wider implementation of the common architecture;

(c) The standards such as Data Documentation Initiative (DDI) and the Statistical Data and Metadata eXchange (SDMX) are complementary rather than competing. SDMX is mainly considered appropriate for transfers of aggregate data, while DDI is more suitable for micro-data. So far, national statistical organisations have

prioritized micro-data, and hence DDI. OECD is developing SDMX based dissemination tools that could be shared;

(d) A more sustainable funding mechanism is needed to support HLG projects and ensure maintenance of outputs. The Secretariat is considering different approaches, including a subscription-based model.

Conclusions

36. The Bureau strongly supported the HLG work, and appreciated the outcomes of the projects in 2013. The Bureau emphasised the need for continuity supported by an appropriate funding mechanism, and the need for HLG to coordinate with other international activities.

IV. Preparation for the Conference of European Statisticians 2014 plenary session

A. Organisation of the Conference of European Statisticians 2014 seminar "Migration statistics"

Documentation: ECE/CES/BUR/2014/JAN/16 by Mexico, Norway, Russian Federation, CIS-STAT, UNSD and UNECE

37. The Bureau discussed the organization of the CES 2014 seminar on migration statistics based on an outline prepared by the seminar and session organizers. The following issues were raised in the discussion:

(a) Countries should be encouraged to focus on issues that are of general interest instead of reporting national experiences;

(b) The session organizers should draft an issue paper to highlight the challenges presented in the papers and to focus the seminar. The questions to be considered during the general discussion should be provided well in advance of the seminar.

Conclusions

38. The Bureau agreed with the proposed agenda and format of the seminar. The session organisers should prepare in advance a short issue paper to focus the seminar and sessions, and a set of questions for general discussion.

39. The seminar will take place on Thursday, 10 April 2014, from 09:30 to 16:10. The seminar organisers will proceed with the preparations, taking into account the recommendations by the Bureau.

B. Organisation of the Conference of European Statisticians 2014 seminar "What is the value of official statistics and how do we communicate that value?"

Documentation: ECE/CES/BUR/2014/JAN/17 by United Kingdom, Eurostat and Sweden

40. The Bureau discussed the organization of the CES 2014 seminar on the value of official statistics based on an outline prepared by the seminar and session organizers. The following issues were raised in the discussion:

(a) The Bureau welcomed the proposed keynote speaker, John Pullinger, President of the United Kingdom Royal Statistical Society and the Librarian of the House of Commons;

(b) Special technological and interactive features (e.g. webstreaming, live questions from the web, vote pads, etc.) should be used only if they add value to the seminar and do not disrupt the discussion. The new approaches have to be tested before using them at the CES;

(c) As the topic is about communication, it is important to keep the debate open and invite stakeholders to join in. This could be done by using low-cost means, such as a video presenting views of young people and the private sector on the value of official statistics;

(d) It would be useful to provide an opportunity for the staff of national statistical offices to remotely follow the discussions;

(e) Mexico will run a survey about the opinions of different groups of users (households, private sector, academia and media) on official statistics and will share the results.

Conclusions

41. The seminar organiser will prepare a list of possible technological services that may be needed during the seminar, ask OECD to provide a quote for these and, if considered appropriate, investigate the possibilities of financing the technical support. The Bureau supported the suggestion to prepare a video to gather the views of young people and private sector.

42. The seminar will take place on Wednesday, 9 April 2014, from 09:50 to 17:30. The seminar organisers will proceed with the preparations, taking into account the recommendations by the Bureau.

C. Recommendations on climate change related statistics

Documentation: ECE/CES/BUR/2014/JAN/18 by the Task Force

43. The Bureau considered whether the draft Recommendations on climate change related statistics are ready to be circulated to all CES members for electronic consultation with a plan to present them to the CES 2014 plenary session for endorsement. The following issues were raised in the discussion:

(a) The Bureau underlined the need for national statistical offices to start implementing the Recommendations, in particular to improve the coherence of greenhouse gas inventory data with other statistics and develop capacities in geospatial information systems;

(b) Turkey expressed its interest in participating in the Steering Group indicated in the draft *Recommendations on Climate Change related Statistics*;

(c) One of the achievements of the Task Force was to establish a close dialogue with the international climate change community. This momentum should be used for further work in collaboration with the climate change community after a consultation of national statistical offices' views and priorities.

Conclusions

44. The Secretariat will circulate the *Recommendations on climate change related statistics* to all CES members for electronic consultation in view of presenting them to the

CES 2014 plenary session for endorsement. The Secretariat will send the Recommendations for comments to the international organisations that have been involved in the work.

45. The Secretariat will consult with the CES member countries proposals for possible future work and will report back to the Bureau in October 2014.

D. Draft agenda and timetable for the 2014 plenary session

Documentation: ECE/CES/BUR/2014/JAN/19 by UNECE

Conclusions

46. The Bureau approved the provisional agenda and tentative timetable for the CES 2014 plenary session (9-11 a.m. April 2014 in Paris).

V. United Nations Economic Commission for Europe Programme of work documents

Documentation: ECE/CES/BUR/2014/JAN/20 and 21 by UNECE

Conclusions

47. The Bureau adopted the report on the implementation of the Statistical Programme 2013 and the Biennial Evaluation Report 2012-2013 (documents ECE/CES/BUR/2014/JAN/20 and 21).

48. The Bureau noted that the Biennial Evaluation Report is not useful for evaluating the CES work, and should be discontinued if possible.

49. The Secretariat will make a proposal to the Bureau meeting in October 2014 on how to evaluate the impact of the CES activities and the various guidelines and recommendations developed under the Conference.

VI. Organisation of Conference of European Statisticians' work

A. Updated Guidelines for organising the Conference of European Statisticians' seminars

Documentation: ECE/CES/BUR/2014/JAN/22 by UNECE

Conclusions

50. The Bureau approved the updated *Guidelines for organising the CES seminars*, including the templates for seminar papers and seminar outlines.

B. Governing rules of the Conference of European Statisticians

Documentation: ECE/CES/BUR/2014/JAN/23 by UNECE

Conclusions

51. The Bureau reviewed the *Guidelines on procedures and practices for ECE bodies*, adopted by the sixty-fifth session of the Economic Commission for Europe, and agreed that the *Rules governing the work of the Conference and its Bureau* are in line with the new

ECE Guidelines. The forthcoming CES plenary session will be informed about the updated UNECE Guidelines.

52. The Secretariat will clarify some recommendations in the Guidelines concerning registration of participants and circulation of the draft conclusions from the meetings.

VII. Presentation of International Labour Organization's activities related to statistics

53. The International Labour Organization (ILO) presented its statistical activities. The ILO statistical work focuses on implementing and promoting resolutions and recommendations of the International Conference of Labour Statisticians (ICLS).

54. ICLS held its 19th Conference in 2013 which marked the 90th Anniversary of ICLS. The 19th Conference adopted, among others, resolutions on statistics of work, employment and labour underutilization; forced labour; and labour migration. The Russian translation of the Resolutions will become available shortly.

55. The work deals with many topics related to labour statistics: measuring wages, labour migration, labour markets, conditions of work, skills mismatch, labour costs and productivity, social protection, child labour, forced labour and human trafficking; rural/urban employment; working time and time-use surveys; statistics on cooperatives; social dialogue statistics; gender mainstreaming in labour statistics; work-related violence; sectoral employment; decent work; and statistics on green jobs. It also relates to the Millennium Development Goals (MDG) and the Post-2015 Development Agenda.

56. ILO maintains statistical classifications: the International Standard Classification of Occupations (ISCO) and International Classification by Status in Employment (ICSE). ILO recently published, in collaboration with partner organisations, three manuals on informality; volunteer work; and employment in tourism sector.

57. ILO works in collaboration with CES in several areas, including population and housing censuses, consumer prices and quality of employment.

58. The following issues were raised in the discussion:

- (a) There are many complementarities and synergies in ILO and CES work;
- (b) CES and ILO could promote closer cooperation among countries with common migration flows to improve the quality of migration statistics;
- (c) Concerns were expressed about the different concepts applied in labour statistics by Eurostat and ILO. The organizations are gradually moving towards common definitions. Green jobs is a new area where international organizations are working together to define common definitions;
- (d) OECD plans to operationalize the concepts developed by the joint UNECE/Eurostat/ ILO/OECD expert group on quality of employment focusing on earnings, labour market security and quality of labour markets. Data collection will start shortly in collaboration with Eurostat and the EU Directorate General on Employment, Social affairs and Inclusion.

Conclusions

59. The Bureau thanked Mr. Diez de Medina, the Director of Statistics Department of ILO for the interesting and informative presentation, and invited ILO to take part in the Conference of European Statisticians on 9-11 April 2014, in particular at the CES seminar on migration statistics on 10 April 2014.

VIII. Dates and venue of the next meetings

60. The CES Bureau confirmed the dates for its next meetings as follows:

(a) A short Bureau meeting will take place during the CES 2014 plenary session to decide on the topics for the 2015 CES seminars and any other urgent matters (a meeting of 15-20 minutes during 9 April 2014 lunch);

(b) The next Bureau meeting will take place on 21-22 October 2014 in Ottawa, Canada;

(c) The subsequent Bureau meetings are planned to be held on 17-18 February 2015 in Geneva and on 20-21 October 2015, venue to be decided;

(d) The Bureau meetings in 2016 are tentatively scheduled to the week of 8 February and 18-19 October, venues to be decided.

61. The 62nd CES plenary session will take place in Paris, back-to-back with the OECD Statistics Committee meeting, starting at 9:30 on 9 April and finishing by 12:30 on 11 April 2014.

62. The 63rd CES plenary session will take place during the week of 15 June 2015 in Geneva. The 64th CES plenary session is tentatively scheduled to be held on 27-29 April 2016 in Paris.



Economic and Social Council

Distr.: General
6 March 2014

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-second plenary session

Paris, 9-11 April 2014

Item 8 (a) of the provisional agenda

Reports on the work of the Conference of European Statisticians, its Bureau and Teams of Specialists

Report of the work session on statistical data confidentiality

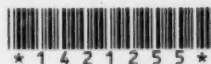
Note by the Secretariat

Summary

This document presents the outcomes of the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, held in Ottawa, Canada, on 28-30 October 2013. The main objectives of the work session were to facilitate the exchange of experience and identify best practices in dealing with technical issues related to statistical data confidentiality in national statistical offices. The meeting targeted experts of national and international statistical offices as well as invited academics dealing with statistical disclosure limitation.

This document is submitted to the Conference of European Statisticians for information.

GE.14-21255



★ 1 4 2 1 2 5 5 ★

Please recycle A recycling symbol consisting of three chasing arrows forming a triangle.



I. Participation

1. The Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality was held in Ottawa, Canada, from 28 to 30 October 2013. It was attended by participants from: Australia, Canada, Finland, France, Germany, Japan, Mexico, Netherlands, Norway, Republic of Korea, Russian Federation, Slovenia, Sweden, United Arab Emirates, United Kingdom and United States of America. The European Commission was represented by Eurostat. Representatives of the World Bank and International Monetary Fund (IMF) also attended. Participants from numerous universities and research institutes attended the work session at the invitation of the UNECE secretariat.

II. Organization of the meeting

2. The agenda of the work session consisted of the following substantive topics:
 - (a) New methods for protection of tabular data or for other types of results from table and analysis servers;
 - (b) New methods for protection of microdata;
 - (c) Modes of access to microdata;
 - (d) The trade-off between quality, utility and privacy;
 - (e) Confidentiality issues and case studies;
 - (f) Panel discussion on transparency and/or privacy concepts (such as differential privacy);
 - (g) Software developments and demonstrations.
3. Mr. P-P. de Wolf (Netherlands) was elected as Chairman. He expressed his gratitude to Statistics Canada for hosting the meeting in Ottawa.
4. Ms. R. Bender, Assistant Chief Statistician, Analytical Studies, Methodology and Statistical Infrastructure Field at Statistics Canada opened the meeting and welcomed participants to the 8th work session in this field. She was pleased to see the large number of countries participating which included the remote participation of the U.S. Bureau for Labor Statistics via WebEx. She noted the continuous changes in technology, data availability and means of communication, which change how we address data issues. Balancing data access and confidentiality remains critical, and methods must evolve to meet new challenges. Academic researchers are key players in the field of confidentiality, and this work session is an ideal opportunity for researchers and national statistical institutes to exchange information and ideas. She wished participants a very good and fruitful work session.
5. The provisional agenda was adopted.
6. The following persons acted as Session Organizers/Discussants: Topic (i) – Ms. S. Giessing (Germany); Topic (ii) – Ms. M. Simard (Canada); Topic (iii) – Mr. P-P. de Wolf (Netherlands); Topic (iv) – Mr. L. H. Cox (United States of America); Topic (v) – Mr. E. Schulte Nordholt (Netherlands); Topic (vi) – Mr. J. Domingo Ferrer (University Rovira i Virgili, Spain); Topic (vii) – Mr. P-P. de Wolf (Netherlands).

III. Recommendations for future work

7. The participants reviewed the recommendations for future work on the basis of a proposal put forward by an ad hoc working group composed of Ms. A. Nissenen (Finland), Ms. S. Giessing, and Mr. J. Drechsler (Germany).

8. The participants considered it useful to continue the exchange of experiences in the field of statistical data confidentiality, and recommended that a future work session on statistical data confidentiality be convened in 2015. Statistics Finland offered to host that work session in Helsinki. The following topics were proposed:

- (a) Data access on an international level;
- (b) Output checking: pros and cons;
- (c) Remote access;
- (d) Remote analysis and table servers;
- (e) New methods for tabular data / microdata;
- (f) Case studies and software demonstrations;
- (g) Risk and utility measures;
- (h) Consistent data release;
- (i) Confidentiality as follows;
 - (i) in the era of big data;
 - (ii) for social networks;
 - (iii) for geospatial data;
 - (iv) for linked data;
 - (v) across different forms of data release;
- (j) Transparency;
- (k) Privacy models considering data utility.

IV. Further information

9. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. All background documents, presentations and the final report for the meeting are available on the website of the UNECE Statistical Division: www.unece.org/stats/documents/2013.10.confidentiality.html

10. On behalf of the participants, Mr. S. Vale (UNECE) expressed his great appreciation to Statistics Canada for hosting this meeting and providing excellent facilities for the work.

V. Adoption of the report

11. The participants adopted the present report before the Work Session adjourned.

Annex

Summary of discussions and main conclusions

A. Topic (i) New methods for protection of tabular data or for other types of results from table and analysis servers

1. Session A

1. The Australian Bureau of Statistics (ABS) has recently developed the TableBuilder and DataAnalyser remote server systems with automated confidentiality routines that allow users to build their own custom tables or undertake for example regression analyses on secured ABS microdata. The presentation described the statistical methodology behind the perturbation and other protection methods used in these systems. The perturbation routines applied in TableBuilder and DataAnalyser are applied not at the unit record level, as is the case with confidentialised unit record files, but at a level of aggregation relevant to the analysis. This results in lower levels of information loss by tailoring the perturbation both to the type of analysis requested and the nature of the underlying data. DataAnalyser is developed in R, so can be shared, whereas TableBuilder uses proprietary software from Space-Time Research.

2. The representative of the University of Kentucky presented a joint work with Oklahoma State University and CSIRO Mathematics on a General Methodology for Masking Output from Remote Analysis Systems. Remote analysis systems are being considered as an effective approach for providing individuals the ability to perform analysis on data held by statistical agencies and to receive the results of such analysis. One of the major obstacles to the wider implementation of these systems is the lack of a masking mechanism that will ensure that the masked response will be both useful and prevent disclosure. It is also desirable that this mechanism is easy to implement and automate so as to minimize human intervention. The presentation suggested a bootstrap mechanism to satisfy these requirements.

3. The representative of the University of Manchester presented work on the development of a flexible table generating server and demonstrated an application of measuring risk-utility comparing different disclosure control methods for census data. They propose measures for disclosure risk and data utility that are based on Information Theory. For flexible table generating, the server has to measure the disclosure risk in the table, apply the disclosure control method and then reassess the disclosure risk. Methods may be applied either to the underlying data used to generate the tables and/or to the final output table generated from original data. Besides disclosure risk, the server should provide measures of information loss comparing the perturbed table to the original table.

4. The German Federal Statistical Office proposed a flexible rounding strategy for ratios also based on stochastic noise. Traditionally, many statistical agencies protect magnitude tabular establishment data by cell suppression. Typical risk concepts, rules and techniques for cell suppression apply to data of an additive nature, i.e. sums of a quantitative variable. For tables presenting means, ratios or other indicators, it is often considered enough to suppress a cell only if it relates to just a single unit. If, however, cell suppression is replaced by a flexible rounding strategy based on stochastic noise, a perturbation might be applied directly to ratios which would reduce the noise variance of the ratio compared to the naïve approach of taking the ratio of the perturbed data. Encouraging first test results were reported, and observed with real life data from German

Tourism Statistics. However, testing of the methodology is still work in progress, especially to prove the concepts suggested for establishing suitable parameters for the method. Another important issue for future work is to test, or eventually extend the method for special, important types of ratio data, like means, i.e. a magnitude variable divided by a count.

5. The CSIRO Computational Informatics presented a paper about their concerns with protecting confidentiality in statistical analysis outputs from a virtual data centre. The presentation described a proposal for a two-stage process involving dataset preparation by the data custodian before loading the data into the virtual data centre and confidentialisation of the analysis outputs by the researcher on removal from the secure environment. The second stage makes use of a checklist developed to assist researchers. However, it would be essential to provide researchers with training in disclosure control as part of the virtual data centre researcher and project approval process so that they could conduct the output confidentialisation themselves.

2. Session B

6. The representative of Statistics Norway presented their work in confidentiality protection of large data cubes. In 2014 all European Union and associated countries will have to submit 60 hypercubes from their 2011 censuses to Eurostat. All the cubes are frequency counts and in most of them the units are individual people. Each country is responsible for disclosure control of their own cubes according to their own legal definitions and with their own choice of disclosure control method. Norway has decided to use a form of balanced rounding of small counts as the preferred disclosure control method. The next step will be to test more advanced methods for the balancing mechanism, like e.g. balanced sampling or mixed integer linear programming.

7. The representative of the US Census Bureau presented lessons his office learned in building a linear programming system prototype of the modernized cell suppression system for his office. The cell suppression problem is mathematically sophisticated, so the development of the methodology for an application built to solve it poses some interesting design problems. By far the most pressing of the problems is a need for overall speed in producing a solution for a particular set of production tables.

8. Controlled Tabular Adjustment (CTA) is a recent approach to compute the closest safe table to the original data, using some distance. Sensitive cells are adjusted either upwards or downwards (binary decision), and the resulting cells have to be accordingly (and minimally) modified to preserve marginal totals. The binary decisions are modelled as disjunctive constraints, CTA resulting in a difficult mixed integer linear problem. The representative of the Universitat Politècnica de Catalunya presented a variant of CTA without binary variables and a software package they have implemented to solve a resulting four-objective optimization problem.

9. The representative of the Statistical Office of the Republic of Slovenia presented a metadata-driven procedure and application for aggregation and tabular protection, as well as the dilemmas and trade-offs of such an approach. The application will be metadata-driven, meaning that there is one general program code which is then parameterised for the particular survey using (process) metadata. The metadata approach causes some additional work for the first time, but for future iterations the metadata can be re-used. For tabular protection the German SAS-Tool which involves Tau Argus will be implemented and used whenever possible. In cases of tables that are very complex, the metadata-driven application will not be a suitable solution, however it should help to construct the input files for Tau Argus. At the same time precision requirements will be implemented and used as an additional rule for primary sensitivity of aggregated data.

10. During the discussion, the following points were raised:

- (a) It is difficult to fully protect large linked tables such as those required for the Eurostat Census Hub. All countries protect their tables according to their own rules but there can be inconsistencies and differences in European totals. The system will be evaluated next year based on lessons learned;
- (b) Participants discussed the extent to which differential privacy can be seen as a noise addition method;
- (c) There is no single "one size fits all" solution for managing disclosure risk. Solutions should be context and problem driven;
- (d) The challenges of managing secondary disclosure risks in the context of generating multiple tables on the fly;
- (e) The possibilities for greater international collaboration, both in the development of sharable code and tools, and in the specification of additional features that statistical organizations would like to see added to commercial software such as SAS. In this context, the Tau Argus package is being released as open source software;
- (f) Where it is not currently possible to share software tools, it may still be possible to share methods.

B. Topic (ii): New methods for protection of microdata

11. Duke University presented their work in developing a fully Bayesian, joint modelling approach for categorical data based on the nested Dirichlet process to protect respondents' privacy in surveys on household data. It induces a two-level clustering structure in modelling the household dataset, trying to simultaneously cluster household members within households, and borrow information across households that have similar clusters. The model is applied to a subset of the Current Population Survey March 2011 household dataset synthesis. The results demonstrate its abilities to preserve marginal and multivariate distributions of all nominal variables (dependence structures among variables), and within household relationships, such as difference in race, age and education.

12. Université Laval presented the problem of releasing microdata from categorical variables in the form of large confidentialized contingency tables. Several algorithms have been designed for this purpose, most of which are based on the idea of multiple imputation. While some offer some strong guarantee of privacy protection, such as differential privacy, others merely rely on the published observations not corresponding directly to real respondents. They took into account the difficulty of offering a confidentiality guarantee when creating synthetic datasets by sampling from posterior predictive distributions, and proposed a possible general mechanism for releasing differentially-private synthetic datasets using this method.

13. The representative of Universitat Rovira i Virgili proposed a framework to attain k -anonymity, based on micro-aggregation of confidential data that seeks the lowest possible variability for the confidential attributes, thereby maximizing utility. The proposal can be combined with k -anonymity refinements such as l -diversity and t -closeness, hence yielding simultaneous utility and privacy guarantees. ϵ -Differential privacy is another privacy model that is often opposed to k -anonymity like models. k -Anonymity is usually presented as a model that preserves data utility to a good extent but offers only limited privacy guarantees. In contrast, ϵ -differential privacy provides strong privacy guarantees but only limited data utility. For microdata releases, ϵ -differential privacy can be seen as a kind of t -closeness with a specific distance measure. The proposal to minimize the variability of the confidential attributes can therefore also be applied for ϵ -differential privacy.

14. The University of Manchester presented an evaluation study to compare statistical disclosure limitation methods for spatial outliers in microdata jointly undertaken with the Artificial Intelligence Research Institute (IIIA) and the Spanish National Research Council (CSIC), Spain. The test dataset for this evaluation study is based on the transportation products of the 2006-2008 combined public use microdata set from the American Community Survey of the United States. The spatial variables are trajectories defined as vectors of coordinates where the first component is the coordinate of place of residence (origin) and the second component is the coordinate of workplace (destination). Variables based on geographical spatial coordinates are particularly prone to disclosure risks since they can easily be visualized when disseminating statistical information through the use of maps. The first stage of the study is an outlier detection algorithm, which accounts for multivariate data and is robust to deviations from normality assumptions. Once outliers are identified, the second stage of the study is to recommend a targeted disclosure limitation method to confidentialize the outliers. The perturbative methods of record swapping and hot-deck with respect to disclosure risk and data utility were compared.

15. The German Institute for Employment Research and Universitat Rovira i Virgili presented their work on evaluating the potential of differential privacy mechanisms for Census data. The large number of available records coupled with a limited set of only a few (often categorical) variables will ensure that most of the cells defined by cross-classifying the different attributes still contain an ample number of records. Thus, the noise that needs to be added to fulfil the differential privacy requirements might have only minor effects on data quality. To enable the release of detailed geographical information they propose a differentially private procedure based on a micro-aggregation algorithm with a fixed minimal cluster size. They evaluated whether meaningful results can be obtained with this approach using administrative data gathered by the German Federal Employment Agency. Detailed geocoding information has been added to this database recently and plans call for making this valuable source of information available to the scientific community. They expect that the proposed micro-aggregation algorithm will enable them to release detailed geocoding information while offering strong differential privacy guarantees.

16. The University of Minnesota presented a joint work on a tripartite collaborative experiment using a ten per cent household sample from the 2011 census of Ireland to estimate risk, mask the data using controlled shuffling, and assess analytical utility by comparing the masked data against the unprotected source microdata. Controlled shuffling exploits hierarchically ordered coding schemes to protect privacy and enhance utility. With controlled shuffling, the lesson seems to be more detail means less risk and greater utility. Overall, despite substantial perturbations of the masked dataset, they find that data utility is very high and information loss is slight, almost imperceptible even for fairly complex analytical problems. IPUMS-International disseminates more than two hundred integrated, confidentialized census microdata samples to thousands of researchers world-wide at no cost. The number of samples is increasing at the rate of several dozen per year, as the process of integrating metadata and microdata is completed. Protecting the statistical confidentiality and privacy of individuals represented in the microdata is vital for the IPUMS project. For the 2010 round of censuses, even greater protections are required, while researchers are demanding ever higher precision and greater utility.

17. The following points were made during the discussion:

(a) Structural zeros are problematic. Further investigations should be made on treating these cases appropriately while creating synthetic datasets;

(b) Differential privacy is defined as participation disclosure, i.e. whether an individual is present or not in a particular data set. However, if participation is mandatory such as in a population census, differential privacy may still be of interest with regard to ensuring confidentiality of outputs;

(c) How to assess goodness of fit of different methods for specific sub-groups? Researchers could help to determine this, particularly in the context of international research data centres;

(d) There remain various legal issues with international research data centres. The Data without Boundaries project is addressing this point;

(e) It is important to ask what the usages are and who are the users when starting research on confidentiality topics. Often one answer is not suitable for all uses or users;

(f) The user community is changing enormously. We are seeing an explosion in terms of users and usage. The challenge is how to provide the different types of data sets needed, whilst maintaining confidentiality. Different balances between utility and privacy may be needed;

(g) As differential privacy was developed for output perturbation, can it be applied to microdata inputs? However, microdata is increasingly being seen as an output, and interactive queries that cannot always be predicted in advance, mean that a differential privacy approach could be relevant;

(h) There is increasing demand for metadata on confidentiality procedures, and transparency can be seen as part of utility, however, the risk to privacy could be increased. An appropriate balance needs to be found.

C. Topic (iii): Modes of access to microdata

18. Papers presented in this session covered remote access within the European context and national modes of access. Two papers were tabled but not presented: Istat on their experience on releasing multiple microdata files stemming from the same survey, and Statistics New Zealand on their Integrated Data Infrastructure (IDI) which is a new feature for the microdata Statistics New Zealand provides access to.

19. The representatives of the German Federal Statistical Office and the Institute for Employment Research, Germany presented their paper on the need to develop a Safe Centre network in order to enrich European research. On the national level, sophisticated approaches to access microdata have been developed, but transnational access and international comparative research are still not easy to do. Two European projects, the ESSnet "Decentralised and Remote Access to Confidential Data in the ESS" (DARA) and "Data without Boundaries" (DwB) are working on improving transnational microdata access in the field of social science research. At the moment, the most common solution is for researchers to travel to the location where the data are stored and work within a Safe Centre. There is, however, a possibility to combine the advantages of Safe Centres and remote access in a single approach of a transnational secure network of Safe Centres. In that case, researchers can access international data from a Safe Centre in their own country. The presentation highlighted the need for an organizational network of Safe Centres bound together by secure remote access connections.

20. Eurostat presented new legal provisions for access to confidential data at the EU level focusing on the modes of access enabled in the new Commission Regulation 557/2013. The advantages and disadvantages of different access systems were analysed from the point of view of their suitability for cross border access to microdata. Due to various changes in technical, organizational and legal environment, the legal basis for access to confidential data for scientific purposes has recently been revised. The main objectives of the new regulation establishing conditions of access to EU confidential data were: to offer more datasets, to allow for new modes of access, to improve access procedures and to adapt them better to the broader legal framework.

21. The representative of the German Federal Statistical Office presented the ESSnet project "Decentralised and Remote Access to Confidential Data in the ESS" (DARA) to improve access to European microdata for scientific purposes. The aim of DARA is to establish a secure channel from a safe centre within a National Statistical Institute (NSI) to the server at the safe center of Eurostat so that researchers can use confidential EU microdata in their own Member States without travelling to Luxembourg. The ESSnet DARA-project team has defined a concept of technical implementation and safety requirements for a remote access system. The concrete task of participating NSIs is to provide a secure channel to guarantee access for data users to the central node and also to provide service and IT-support for the researchers on the local national level. The project team has drafted a handbook with descriptions and guidelines for NSI staff and researchers and an accreditation system for access facilities. For a proof of concept and feasibility, the project team has implemented a remote access pilot with 5 access points in 4 countries, and a central node in France.

22. The German Institute for Employment Research presented the main components of the proof of concept pilot project for a European Remote Access Network (Eu-RAN) within the Data without Boundaries project funded by the European Commission. Eu-RAN will be built around a single point of access that allows connecting researchers from different locations with data sources located in different European countries via secure remote access solutions. Additional tools like a sophisticated user account management, a Virtual Research Environment (VRE) and a Microdata Computation Centre (MiCoCe) complete the services offered by the Eu-RAN. While the goal of this work package is to build and proof a technical concept, technical aspects cannot be addressed without taking into account legal, organizational and financial issues.

23. The National Institute of Statistics of France (INSEE) presented their policy in allowing researchers to access confidential business data and confidential household data. The various ways of accessing data include specific tables made for researchers and access through a safe data centre. Access for foreign researchers, access to fiscal data when linked to statistical data and the use of output checking in a safe data centre were also discussed. The presenter outlined his personal view that output checking is not effective, because researchers could simply commit to memory some specific numbers, or hide some numbers in a specific text. Output checking is also resource intensive, and would focus the blame on the statistical organisation rather than the researcher if confidential data are disclosed.

24. The presentation by the Netherlands described the various options they provide for access to microdata from a historical point of view, including public use files or microdata under contract. On site access is provided, but remote access is becoming increasingly popular with researchers. In addition, co-operation projects in which Statistics Netherlands works closely together with one or more external partners are becoming more common. Different statistical disclosure control rules are applied for the different access methods, and overall responsibility for microdata access is centralised in one unit.

25. The United Kingdom Data Service introduced and shared some practical issues and experiences encountered with everyday output checking requests. They also suggested how practical/contextual and data specific issues may affect decisions on whether to release an output or not. This in turn may raise more questions than answers given the complex nature of statistical disclosure control. They also discussed the mandatory training of researchers to minimize problems.

26. The following issues were raised during the discussion:

(a) How to ensure consistency in output checking. Some organizations have a policy of double checking of outputs, whilst others have an approach based on standards and guidelines;

(b) Linked data can augment problems, as it is increasingly the case that many sources are used for many outputs, rather than the traditional one-to-one relationship. The 2009 UNECE publication on Principles and Guidelines on Confidentiality Aspects of Data Integration was cited as useful guidance in this area. It was said that the linkage itself is not the cause for additional confidentiality issues. The fact that the linked datasets are enriched however, might increase the risk;

(c) The consistency of metadata relating to microdata sets was raised. Eurostat provide standard quality reports to researchers with the microdata;

(d) Remote access is likely to become the main mode of access in future for the research community;

(e) At the European Union level, legal and technological constraints are diminishing, but resource constraints are growing. Funding models are needed for microdata access services;

(f) "Campus files" should be considered as public use files to be used for educational purposes. Depending on the educational goal, the requirements may be different to "ordinary" public use files. It will be difficult if not impossible however, to produce real public use files with enough scientific content to satisfy researchers;

(g) Remote access may offer lower risks than scientific use files because the raw data do not physically leave the organization;

(h) The European Statistical System develops guidelines and organizes training on output checking, to improve consistency.

D. Topic (iv) The trade-off between quality, utility and privacy

27. The mission of national statistical offices is to release high-quality statistical data, without bias or censorship. However, national statistical offices are also responsible to preserve the confidentiality of data pertaining to individuals collected and used for statistical purposes. This is called disclosure limitation and often involves censoring (suppression, topcoding), modifying (perturbation, tabular adjustment) or replacing (synthesizing) original data prior to releasing or providing access to a final data product. Ideally, the final product both exhibits low to acceptable disclosure risk and is a suitable substitute for original data in terms of usability and outputs from statistical analysis of the released data. These two objectives are often in conflict; resolving this conflict is referred to as balancing data confidentiality and data quality. Six contributions to this Session represented a variety of countries and disclosure limitation methodologies.

28. The representative of the University of Kentucky presented their joint work with Oklahoma State University/University of Alicante on the need for a common set of performance metrics to maintain consistency of output from remote access servers. Government agencies release information in the form of tabular data, microdata, and increasingly in recent years, through remote access servers. The different forms of release are intended to satisfy the needs of different users. While this flexible approach for releasing data is a welcome sign, it also creates new issues that must be addressed carefully. Failure to maintain consistency across the different forms of data release could result in losing user trust in the information that is provided. While there are performance metrics associated with each of the individual forms of data release, they are tailored to the specific form of data release. Adopting common performance metrics is necessary in order to maintain consistency across all forms.

29. Minimum distance controlled tabular adjustment (CTA) is a recent perturbation approach for statistical disclosure control in tabular data. CTA looks for the closest safe

table, using some particular distance. The representative of the Universitat Politècnica de Catalunya, Spain, provided empirical results to assess the disclosure risk of the method and summarized the results reported in the paper. A set of 33 instances from the literature and four different attacker scenarios are considered. The results show that, unless the attacker has good information about the original table, CTA has low disclosure risk.

30. The representative of the French National Institute for Statistics and Economic Studies presented the advantages and disadvantages of three methods of dealing with a secondary cell suppression problem in Tau-Argus in terms of quality and usability, and examined the time needed by algorithms to converge against the loss of information contained in suppressed cells. Typical cases are tested on real data sets: two hierarchical variables, sets of linked tables etc. Some guidelines were then proposed to find the best trade-off between quality of released data and time consumption.

31. The representative of the Artificial Intelligence Research Institute presented their work with Stockholm University on the Hellinger distance for measuring information loss in microdata. Literature presents different approaches to measure information loss in data protection. The probabilistic information loss measure is based on comparison of means, variances, covariance, correlation coefficients and quantiles. In a recent paper, the Hellinger distance has also been used to measure information loss. This distance has been used to compare univariate probability distributions and to measure information loss. A set of experiments using numerical files protected using a few data protection methods was presented.

32. The International Monetary Fund presented the policies, procedures, and IT implementations put in place to safeguard confidential data, and reduce the possibility of accidental disclosure, with specific examples from the ongoing data management procedures within the IMF Statistics Department. The topics covered were: Levels of Data Confidentiality - with respect to the commitments with National Authorities, and internal Fund policies; Procedures - omissions and suppression of data to limit residual disclosure, and safeguard information derived only for the purpose of internal analysis; Technology - data management and dissemination system and user-level restrictions to limit the access to confidential data; Challenges - in terms of balancing transparency of data dissemination and the need to safeguard dissemination of unauthorized data.

33. The representative of the United States National Institute of Statistical Sciences explored the use of circuits for analysis of suppressed tabular data. Theoretical methods and software are available for performing optimal complementary cell suppression (CCS) in tables. The released resulting suppression patterns comprise algebraic circuits which define alternative tables for the original table while controlling variation between original and alternative cell values. For an important class of statistical tables including two-way tables, these circuits are simple alternating (+/-) cycles. Suppressed tables are notoriously difficult to analyse statistically. A user with sufficient resources could construct the full set, a subset or a probability sample of alternative tables and analyse these tables, resulting in bounds for or estimates of analytical outcomes for the original table.

34. The following issues were raised during the discussion:

(a) Methodologists don't necessarily have the power to make decisions on the optimal trade-off between quality, utility and privacy, but should try to influence the decision makers;

(b) We have been creating piecemeal solutions to specific problems for many years. This may not be sustainable. We should look for different and more consistent ways to address confidentiality problems from a wider perspective;

- (c) Lack of consistency can be a significant disclosure risk for some methods. This risk can be lower for perturbation methods;
- (d) When agencies share data, if they apply different disclosure protection methods there can be a real risk of accidental disclosure;
- (e) Why make the user reconstruct the table when we can release one reconstructed table at random? It can achieve consistency and utility at the same time;
- (f) The aim of the researcher is to provide the most useful information and at the same time ensure privacy;
- (g) The need for perturbation in regressions, particularly for units with high leverage.

E. Topic (v): Confidentiality issues and case studies

35. The representative of the Office for National Statistics of the United Kingdom presented their paper describing an "intruder test" where ONS staff were supplied with actual disclosure-controlled 2011 census tables for their local area and under secure conditions, attempted to identify people or households and consequently discover attributes. He summarised the results and showed how they indicated sufficient uncertainty to validate the effectiveness of the statistical disclosure control policy for 2011 UK population and housing census data.

36. The representative of Sweden presented their experiences of implementing statistical disclosure control with the Bifrost application. Their experiences, both positive and negative, were presented and a solution to the problem of integrating consent control into the process of disclosure control was described. To be able to publish tables where information about businesses can be disclosed, Statistics Sweden has to ask the businesses for written consent. Statistics Sweden has recently put a lot of effort into developing standard tools for different parts of the statistics production process. For statistical disclosure control, they have developed "Bifrost", which includes SAS2Argus, a collection of SAS macros that facilitates the use of the Tau-Argus program via SAS. Copies of SAS2Argus are available for anyone that is interested.

37. Statistics Finland produces around 200 different sets of statistics per year, and thus has a variety of statistical disclosure control (SDC) practices in use. In 2012, the office's internal guidelines on the protection of tabulated personal and enterprise data needed updating. Parts of the old guidelines were too open to different interpretations on the necessity and appropriate practices of SDC. There was strong demand for more harmonised methods and practices inside the office, and also opening the black box of SDC to customers and users of statistics. New guidelines were prepared during 2012 and they came into force in February 2013, an English version is available. The next step towards more harmonised SDC is to complete the adoption of the new guidelines in practice.

38. The representative of Meikai University, Japan, presented an approach to assessing the effectiveness of disclosure limitation methods for census microdata. Following the revision of the Statistics Act, anonymized microdata from official statistics have been released in Japan since April 2009. Six types of anonymized microdata from Japanese official statistics are available. In addition, there are plans to release anonymized microdata from the population census. Several empirical studies on the effectiveness of disclosure limitation methods such as micro-aggregation, additive noise, and data swapping for official microdata have been conducted by the National Statistics Center, Japan. The effectiveness of different disclosure limitation methods including non-perturbative methods

and perturbative methods, and their potential for the creation of anonymized official microdata in Japan, were discussed.

39. Statistics Canada presented the two methods used by the Canadian Centre for Justice Statistics (CCJS) to protect tabular data: the scoring approach and additive controlled rounding. The disclosure method for tables uses a scoring approach whereby all survey variables are assigned a sensitivity score. A table's score is the sum of the scores of the contributing variables. If the score falls above a threshold value, then the table cannot be released. This approach is applied to tables that CCJS produces for its own publications as well as for custom requests that it receives from external clients. CCJS is also making data available through the Real Time Remote Access (RTRA) tool which allows individuals to submit SAS code to produce tables. The RTRA uses controlled rounding to mask sensitive data.

40. The session chair summarised the key points raised:

- (a) Linking statistical software packages can improve efficiency, and could be a topic for future meetings;
- (b) Asking consent to disclose information – more experiences would be interesting;
- (c) Sharing experiences on the implementation of new confidentiality rules – are these rules known by producers and accepted by users?;
- (d) Protection methods – how effective are they in terms of balancing risk and utility?;
- (e) How to manage statistical disclosure for very sensitive data – can such microdata be released?

41. During the discussion, the following issues were raised:

- (a) Maintaining a record of data releases to see if new requests are close to previous ones;
- (b) The application of intruder testing to census microdata – some experiments are under way in the United Kingdom;
- (c) Noise in data can disrupt correct identification of individuals, which could be an argument for lower levels of aggregation in outputs, e.g. single year age bands rather than 5-year bands;
- (d) Disclosure control guidelines need to be very precise and mandatory to avoid diverging interpretations. Different guidelines are needed for personal and business data;
- (e) Documenting different disclosure control practices can be a useful step towards harmonisation, as it provides a basis for discussion;
- (f) Perceived disclosure can be as harmful as real disclosure, particularly for sensitive data;
- (g) Eurostat has recently completed a survey on disclosure control tools and methods in member countries. In most countries the application of tools and methods is the responsibility of the subject-matter units. Clear guidance and simple procedures are therefore important;
- (h) The possibilities for international collaboration in the development of tools and methods, for example based on the SAS2Argus macros used in several countries.

F. Topic (vi): Panel discussion on transparency and/or privacy concepts (such as differential privacy)

42. Different views on the issue of transparency versus privacy were presented by the panellists:

(a) As discussed in the panel, transparency refers to the information about SDC methods applied to protect the data; transparency should also be ensured to explain microdata access conditions and requirements;

(b) From the users (researchers) point of view transparency helps to interpret the data correctly. If the information about SDC methods is not provided to the users, a wrong statistical conclusion can be drawn. If the information about SDC method is transparent the users can:

(i) Evaluate the limitations of the released information;

(ii) Correct for the SDC mechanism in the analysis phase;

(iii) The risk of re-identification of statistical unit is very low; moreover, there are not known cases of malicious identification; the known cases are rather due to lack of data security;

(c) From the statisticians' and data owners' point of view:

(i) Transparency may lead to additional disclosure risk; the SDC protection may be redone if parameters of the SDC methods are known to the users;

(ii) In the process of SDC protection of the data statisticians should ensure that the SDC does not change drastically the structure of the data and that data utility is preserved as much as possible;

(iii) There are technological solutions that should help to find out who has breached statistical confidentiality; provision of more data in a more transparent way should be considered ensuring at the same time that confidentiality breaches are appropriately sanctioned;

(iv) The few (zero?) malicious cases of disclosure are due to adequate SDC protection; if the SDC protection were reduced there would be more confidentiality breaches.

43. J. Bambauer – Tragedy of the De-Identified Data Commons: An appeal for transparency and access. Most policy discussions about research data conclude that the privacy risks are too great, and that research data cannot be released safely. The presentation challenged the dominant wisdom, arguing that properly de-identified data is not only safe, but of unmatched social utility claiming that:

(a) First, many people have misinterpreted the relevant literature from computer science, and thus have significantly overstated the difficulty of anonymizing data;

(b) Second, the available evidence demonstrates that the risks from anonymized data are hypothetical - they rarely, if ever, materialize;

(c) Third, the alternative approach taken with Differential Privacy has extremely limited application. Outside a narrow range of circumstances, it fails to deliver either privacy or utility or both;

(d) Finally, de-identified data is crucial to beneficial social research, and constitutes a public resource - a commons - under threat of depletion.

44. A-S. Charest – Statistical Need for Transparency: Statistical Disclosure Control (SDC) methods always involve a trade-off between privacy protection and data accuracy. I will argue that transparency of the SDC mechanism should be an important part of the data accuracy evaluation. It is necessary not only to provide the analyst with data that is as accurate as possible but also with an appropriate measure of the accuracy of this data. Ignoring the additional variability introduced for the purpose of confidentiality may lead to inaccurate inference. Agencies should thus either provide data which can be correctly analysed in the format in which they are given, or provide enough information for users to take the mechanism into account in their analysis. Examples of both approaches were discussed.

45. L. Cox – Transparency Issues for Tabular Data: Statistical disclosure limitation (SDL) is necessary to protect the confidentiality of information pertaining to subjects of surveys and other statistical inquiries. SDL methods typically remove, modify or replace original data, thus distorting data, total mean square error and, potentially, inference and analytical outputs such as regression coefficients. Transparency in SDL refers to equipping analysts with information about the SDL method and its effects on statistical analysis--provided that this information does not lead to inferential disclosure. Thus, for example, if the SDL introduces bias, inflates variance or attenuates estimated regression coefficients, then the analyst may be provided with estimates of bias, inflation or deflation for the purpose of correcting computed estimates. (Nearly) everyone would agree that transparency is a good thing, but like truth and beauty, can be difficult to assess, obtain and preserve. This is true for tabular data subject to suppression as these data involve complicated relationships between true and alternative values that in some situations may be deconstructed or even undone.

46. During the discussion, the following points were made:

- (a) Transparency can therefore be considered as an opportunity (as it increases data utility) and a danger (as it increases the disclosure risk);
- (b) The catalogue of SDC methods assigned with the risks related to the transparency could be helpful in taking the decision how much information could be provided to the user;
- (c) The information available to the user should be taken into account when evaluating the risk of disclosure;
- (d) Particular values of parameters of the SDC methods should be rather kept secret, only general protection methods should be transparent;
- (e) In the process of protecting the data the statisticians should ensure that correct inferential analysis can be performed with the protected data;
- (f) Effects of the SDC methods on utility should be measured and adequately balanced with the level of protection.

G. Topic (vii): Software developments and demonstrations

47. This session provided an opportunity for organizations to show recent developments in their software. The following presentations and demonstrations were made:

- (a) Open source software Argus (Netherlands) – The “Argus-twins” are two packages that have been developed within several European projects (funded by Eurostat). The maintenance of the software has been in the hands of a small group of people, mainly at Statistics Netherlands. Since one of them has reached the age to retire, a shortage of developers/maintainers threatens. Moreover, the Argus software was developed for the

Windows operating system. These two arguments have led to a new project (partly funded by Eurostat) in which the Argus software will be ported to an Open Source version. We will show the ideas behind this porting, discuss the expected advantages and show some preliminary results;

(b) G-Confid: Turning the tables on disclosure risk (Canada) – New business survey confidentiality software G-Confid, created by Statistics Canada was based on methodology used in the old Confid system for cell suppression in tabular data. It can handle any table size & number of dimensions subject to SAS & hardware limitations. This software is user-friendly, has the same look and feel as Statistics Canada's other generalized systems and can incorporate new approaches. The main objective of G-Confid is to provide the appropriate level of protection for confidential cells while minimizing the loss of information. To achieve this objective, G-Confid uses an automated linear programming procedure to optimize residual suppression. The presentation covered the functionality and characteristics of G-Confid, focusing on new features including the MINRESP and SCALECOST functions, as well as new auxiliary functions for analysts;

(c) G-Tab: Generalized Tabulation Tool (Canada) – In the social statistics field in Canada, there is no common process for disseminating tabular information. Standards for table production, confidentiality and quality vary. Several tools already exist to manage the table creation process, while Statistics Canada's new dissemination strategy is putting pressure on divisions to disseminate more information to the public through CANSIM and other on-line tools. The increase in tabular information being offered to the public increases the risk of potential breaches of privacy. Due to the complex nature of many of the surveys, the creation of estimates and quality indicators is complicated and prone to human errors which affect quality. G-Tab resolves the inconsistencies in standards in applying confidentiality and updating confidentiality rules to enable better protection in the changing dissemination environment at Statistics Canada. G-Tab has a fully automated set of confidentiality rules that will help protect the information published via free CANSIM;

(d) Innovative Microdata Access - Confidentialising on the fly (Australia) – The Australian Bureau of Statistics (ABS) has recognised the need to 'rethink' microdata access to meet increasing user demand for access to more detailed unit record data, across a wider array of datasets. Until now, ABS data access strategies have relied predominantly upon the confidentialisation of the microdata prior to their release. ABS has recognised that this approach is not sustainable, due to the considerable burden it places on resources and the protracted length of time to release. In addition, this one-size-fits-all confidentialisation solution has to cover all possible users and all possible analytical uses of a dataset, which therefore results in additional loss of information for users. The ABS has developed a remote execution environment for microdata consisting of two different but complementary services; Survey Table Builder (STB) and Analysis Service (for statistical analysis). STB includes an innovative approach with a routine that dynamically confidentialises tabular output using perturbation techniques prior to it being returned to the user. The primary advantage of this technique is that the impact on overall data quality of the confidentialisation routine is minimised by applying the necessary modifications only to the final output and not to the underlying data source. The Analysis Service functionality includes data transformations and manipulations, summary tables, exploratory data analysis and regression models with graphical and diagnostic outputs. As for the STB, all outputs are confidentialised 'on the fly' to protect data confidentiality, using perturbation methods and other protections.